DOCUMENT RESUME

ED 169 414 CG 013 348

AUTHOR Fine, Michelle; Saxe, Leonard
TITLE Evaluation Research and Psychology: Toward Synthesis.
PUB DATE Aug 78
NOTE 16p.; Paper presented at the Annual Convention of the American Psychological Association (86th, Toronto, Ontario, Canada, August, 1978)
AVAILABLE FROM Department of Psychology, Boston University, 64 Cummington St., Boston, Mass. 02215

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Evaluators; *Information Processing; *Program Evaluation; *Program Validation; *Psychological Studies; *Research Methodology; *Systems Approach

ABSTRACT

The extent to which evaluation research and psychology share some common objectives, methods and aims for validity is examined. The conventional dichotomy drawn between the two disciplines is seen as spurious and misleading to scholars and practitioners in each area. To demonstrate the proposed synergy between the two disciplines, an integrative model of research is presented to which a dual-level methodology combines the methods and results of traditional evaluative research and experimental social psychology. This macro-micro integrative model of research facilitates systematic program evaluation toward three ends: (1) information is generated for further theoretical understanding of the value and validity of the psychological construct underlying the program; (2) information on the operational level indicates how the theoretical construct works in practice; and (3) the on-line implementation of the construct is assessed through systematic investigation which separates construct and implementation. The dual-level methodology, in which ongoing time series data complement information generated in traditional experiments, provides information relevant to the specific situation as well as information generalizable beyond the boundaries of the program. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED169414

Evaluation Research and Psychology: Toward Synthesis*

Michelle Fine
Teachers College, Columbia University

Leonard Saxe
Boston University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M Fine

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM"

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

*This paper was presented at the Meetings of the American Psychological Association. Toronto, August, 1978. Comments and requests for reprints should be addressed to Leonard Saxe, Department of Psychology, Boston University, 64 Cummington Street, Boston, MA 02215.

CG013344

Rather than approach evaluation research and psychological research as separate activities that should be carefully distinguished from one another, our comments rest on the assumption that there is a basic parallelism between evaluative and psychological research endeavors (Bonoma, 1976; Mitroff & Bonoma, 1978; Wortman, 1976). Such parallelism should not be surprising, given the involvement and prominence of psychologists in the development of evaluation research as a discipline (cf. Perloff, Perloff & Sussna, 1975) and the fact that the methods of evaluation research borrow heavily from psychology (see Guttentag & Struening, 1975). Our view is that the nature of this parallelism needs to be examined. In order to do so, the present paper will analyze the metatheoretical and methodological features of these two modes of social science inquiry, to demonstrate how the better integration of evaluative and psychological research can lead to an intellectually sound and socially useful model for studying human behavior.

Perhaps, by way of establishing the context for this discussion, what we mean by evaluation research and psychological research should be specified. Evaluation research is characterized, traditionally, as a means to an end--a method to provide information about social programs that can be utilized for decision making purposes in a particular context or series of contexts (Franklin & Thrasher, 1976; Suchman, 1967; Weiss, 1972). Psychological research, in contrast, is usually defined as an end in itself--the use of scientific methods in the development of a psychological theory of human behavior (e.g., Schlenker, 1974). Kruglanski (1975), commenting on the validity of laboratory research, has introduced two terms which neatly capture how evaluation and

psychology are typically distinguished. He uses the term "particularistic" to identify research which is aimed at understanding particular situations and instances of behavior. The traditionally assumed function of evaluation research, to provide information about the efficacy of a particular program, fits the particularistic definition. In contrast, Kruglanski uses the term "universalistic" to describe research which is not bound to any particular situation, but instead focuses on the detection of invariances or patterns in behavior across situations. Universalistic research is the ideal ascribed to by most people engaged in psychological research.

Our position is that this dichotomy, as it is applied to evaluation research vs. psychological research, is false, based on spurious assumptions about both modes of research. Evaluation research, while appropriate to decision making, isn't necessarily limited to particularistic research. While evaluation research may typically be conducted to assess the effects of a specific social program, it is frequently designed to provide data for the development of broad scale social policy. The program studied is often but an operationalization of a class of programs about which one wants to generalize.

In considering the major Federal evaluations conducted in the last decade, the non-particularistic, in fact, universalistic character of evaluation should be apparent. The ill-fated evaluation of Head Start (Cicerelli, et al., 1969; see, also, McDill, McDill & Sprehe, 1972, Williams & Evans, 1969), the New Jersey Negative Income Tax Experiment (Kershaw, 1972; Rossi & Lyall, 1976), and the current Rand National Health Insurance Experiment (see Newhouse, 1974) are but a few examples of program evaluations that were conducted primarily to develop gener-

alizable findings. Similarly, smaller scale evaluation studies, such as the Kansas City Police Patrol Experiment (Kelling, Pate, Dieckman, & Brown, 1974) and the Cambridge-Somerville Youth Experiment (see McCord, 1978) were designed to have implications beyond the scope of the actual program.

While evaluation research is probably more universalistic than it is usually given credit for, traditional psychological research is also probably more particularistic than is usually admitted. There is little need for us to elaborate on the artifacts which threaten the validity of laboratory findings, as many commentators have elsewhere discussed the artificial features of the psychological laboratory (see, for example, Rosenthal & Rosnow, 1969). While it is probably too severe to suggest that psychological research is invalid, because of these features of the laboratory, it is probably fair to note that such research may be rather particularistic (e.g., limited to situations where college students interact with strangers).

These comments about generalizability are prelude to an important issue--the use of research (or lack of same) in theorizing about complex human behavior. Here, too, we think an artificial distinction has been made between evaluative and psychological research, in which the former is considered atheoretical, while the latter is considered the tool of the theorist. Evaluative research, directed at the assessment of social programs, is dedicated, primarily, to the development and improvement of our understandings of social functioning. Such research findings, therefore, have critical implications for our theoretical models of human behavior. Social programs are not random attempts to foster social change. Even our most naive social programs, rooted in loose

conceptualizations, reflect some a priori predictions of which program variable will elicit which type of participant response--evidencing some theoretical model.

We would like to suggest, therefore, that not only do social programs lend themselves to the testing of theoretical hypotheses about participant behavior, but further, some lines of theoretical inquiry may best be conducted in such contexts. For example, for research on diffusion of innovation, technological habituation or sleeper effects in the real-world, the long termedness of evaluation studies is ideal. In fact, the evaluation of social programs affords us the opportunity to respond to criticisms of the use of acute measures in research--that psychologists cannot because of the "acute-ness" of our measures, demonstrate process issues, and that psychologists may present a distorted "reality" by measuring only immediate effects (McGuire, 1967). For example, it has been proposed that in the famed Stanford prison experiment only adaptation or transition data were actually collected, rather than data reflective of realistic prison-life circumstances (Banuazizi & Movahedi, 1975). Evaluation research of an authentic prison program conducted over and extended period of time, could collect data more comprehensive or "realistic". The constraints of the laboratory often make acute measures the only option available to psychological researchers. Such measures present risks in so far as they may preclude accuracy and validity in data collection, obscure conceptually critical inconsistencies or change over time, and resulting in meaningless or deceptive data. Evaluative contexts, in contrast to one-shot laboratory experiments, lend themselves to long-term investigation--and may be a most appropriate site for theoretically studying psychological processes

over time.

Up to this point, we have focused essentially on the goals of evaluative and psychological research—identifying areas of commonality and divergence. Let us now address the methods of research. It may come as no surprise that the logic of inquiry, the approaches to studying problems employed in evaluation and psychological research are, indeed, quite similar. In fact, the nature of methodological controversy is even common to the two disciplines. To begin, both disciplines are plagued by the all too common intra-disciplinary debate over the "appropriateness" of particular methodologies. There is considerable confusion, both in the psychological research literature and the evaluative research literature about the "rightness" of particular methodologies for the respective modes of inquiry. Yet, in both disciplines, methodological rigidity is rapidly being replaced by scientific elasticity and flexibility. While some psychologists have posited that only experimental approaches produce meaningful psychological data, a variety of alternative research strategies are now being adopted (e.g., Gergen, 1977). While some evaluation researchers (e.g., Guttentag, 1975) have argued that non-experimental paradigms are the only feasible approaches to evaluation research, there has been a great deal of counter comment in support of true social experiments (cf. Boruch, 1975; Cook & Campbell, 1976). So, although the self-concepts of evaluation and psychology methods may differ significantly, their polar positions are eroding, with each moving toward adoption of a repertoire of common methods.

At the ultimate intersection of methods of evaluation research and psychological research sits the "social experiment". Recent discussions of the development of an "experimenting society" (Campbell, 1969, 1971)

suggest an important unity between the methodologies and theoretical interests of evaluative and psychological research. In the experimenting society, evaluative tests of the efficacy of social programs and theories are conducted as true experiments. The inclusion of true experiments in the process of policy development represents the extension of the basic research model of psychology to broader real-world contexts.

How does an experiment, conducted as part of social evaluation, differ from an experiment conducted as part of a psychological research endeavor? The answer lies in the nature of the independent and dependent variables used in the experiment. In most instances, program evaluation studies utilize more global and complex independent variables than those used in traditional psychology. Probably one of psychology's most obvious failings is its inadequate appreciation and treatment of complex human behavior. Particularly, within social psychology, there have been critiques (e.g., McGuire, 1973) of psychology's emphasis on univariate and bivariate relationships, to the exclusion of more complex (but, naturally occuring) relationships. Evaluative research presents an opportunity to study, scientifically, such intricate and complex relationships. If psychological researchers were confronted with evaluation-type research problems, psychological theories would be forced to open up, to account for new and novel factors, for complex independent variables which influence human behavior. The price we pay is sacrificing our routinized use of simple univariate or bivariate models (à la Aaron, 1978). The benefits include the development of theoretical models with construct validity and real-world applicability.

To the extent that social experiments complicate our lives with

entangled independent variables, they offer us rich diversity on the dependent variable side. The conduct of social experiments supplies us with the opportunity to view critical variables in process, in transition, and in interaction with other variables. Although psychologists do conduct long term studies, particularly within developmental psychology, they are all too few, cumbersome to implement and difficult to complete. The conduct of social experiments provides a context for psychological researchers to study change over time--a context in which documenting change is not only facilitated, but required. Discrete or continuous measures of change concur perfectly with the conduct of evaluative research and, as noted, can benefit our theoretical speculations about particular social and psychological processes. This model of research affords us the chance to collect a wealth of information at different moments in the history of a social program. Data on identical, as well as different, intervening and dependent measures can be collected throughout the course of the research. Researchers who employ this model are not limited to one or two dependent measures but, instead, may enjoy the luxury of limitless measures of program impact.

Despite the extensive cross-discipline overlap of evaluation and psychological research, it is clear that the degree of theorizing and conceptual rigor distinguish the two modes of scientific inquiry. Evaluators, for the most part, do not spend a great deal of time carefully working out a model of the processes they are studying, nor do they carefully analyze their expectations of results; that is, lay out hypotheses. In many cases, while evaluators act as methodological consultants, and program administrators are primarily concerned with delivery of services, a theoretical vacuum results. In contrast,

psychological researchers may focus on the elaborate specification of assumptions and theorizing at the risk of employing trivial or insignificant operationalizations of their concepts.

Our analysis suggests that the synthesis of activities of evaluative researchers and psychological researchers, an important integration, is both possible and practical. Combining the methodological and operational sophistication of evaluation research with the theoretical sophistication of psychology should make it possible to improve the conduct of both forms of inquiry. One mechanism through which this synthesis may be achieved is to conduct, concommitantly and complementarily, evaluative and basic research studies. Such a synthesis should greatly enhance our ability to scientifically understand the processes of complex social behaviors that are site-specific and generalizable.

To facilitate this synthesis, we have developed a model of research (see also, Saxe & Fine, 1977) which generates data relevant to the advancement of social theory and social policy: a dual level methodological approach comprised of what we call a macro level assessment and a complementary micro level assessment--each level appropriate to generating a distinct type of data. The macro level involves a time series design to monitor the ongoing nature of a social program, from inception through termination. The macro level reflects, most simply, traditional evaluative research. Pre, during and post implementation data are collected from participants and non-participants, yielding a profile of programmatic performance which should, if conducted appropriately, indicate the random and systematic fluctuations of program performance. The micro level, alternatively, involves a series of laboratory-like experiments designed to collect data on particular components of the

program--data which are ultimately generalizable, both within and beyond the confines of the actual program being assessed. The micro level parallels pure psychological research to the extent that it employs scientific rigor to promote theoretical advancement.

This model of research facilitates an understanding of three levels of any social program:

(a) the theoretical level--assessing the value and validity of the psychological constructs underlying the social program; a testing of the construct in the abstract.

(b) the operational level--assessing the value and validity of the program operations as reflected in the actual social program; a testing of the construct in practice.

(c) the implementation level--assessing the relative costs and benefits/sucesses and failures of the implementation of the social program; a testing of the administration and implementation of the programs as instituted, rather than as constructs or theoretical entities.

One of our basic assumptions is that the macro-micro level assessments be conducted as part of a dialectical effort to understand the workings of the evaluated program; that is, there be a cyclic exchange of information on from a study of the on-line program and a testing of specific program components. Problems apparent on the macro level are isolated and examined, in detail, on the micro-level. To complete the cycle, findings from the micro-studies instigate programmatic changes, or shifts in emphasis of the macro level investigation. In such an action reserch model (cf. Lewin, 1948), data from the two levels are recycled to identify those aspects of the program that require revision,

expansion, or some other form of special attention. The resulting data provide a conceptual understanding of the reasons for program success and/or failure, while advancing our theoretical ideas about complex human behaviors.

We recognize that the present review has been relatively non-substantive; few comments have addressed the content foci of evaluation research and psychology. We assume, given the present audience, that there is agreement amongst us that the problems of society have important psychological components. The study of human behavior, be it evaluative or a more traditionally psychological study, should inform our choices of public policy. We have argued that only when the problems of psychology are investigated across a variety of contexts (including both laboratory and real world varieties), can we develop and refine our theoretical understandings and ultimately social policy. Further, we have proposed a model of integrative rigorous research which lends itself to real world hypothesis testing about program performance, the adequacy of program components, and psychological theory testing.

Our comments, though they no doubt espouse psychological chauvanism, suggest that although evaluation research is clearly an interdisciplinary endeavor, it does contain a unique psychological contribution; psychologists have been and will be important figures in the field (Wertheimer et al., 1978; Wortman, 1975). Although psychology is capable of studying a wide range of problems through a variety of foci, the interdisciplinary nature of evaluation can't be ignored. To develop a psychology which is relevant and applicable to society's problems requires interdisciplinary collaboration, as well as intradisciplinary thinking. The integration of evaluation within psychology can promote this type of interdisciplinary practice.

References

Aaron, H. J. Politics and the professors: The Great Society in perspective. Washington, DC: The Brookings Institution, 1978.

Banuazizi, A., & Movahedi, S. Interpersonal dynamics in a simulated prison: A methodological analysis. American Psychologist, 1975, 30, 152-161.

Bonoma, T. V. Social psychology and social evaluation. Representative Research in Social Psychology, 1976, 7, 147-159.

Boruch, R. F. On common contentions about randomized field experiments. In R.F. Boruch and H.W. Riecken (Eds.), Experimental testing of public policy: The proceedings of the 1974 Social Science Research Council Conference on Social Experiments. Boulder, CO: Westview Press, 1975.

Campbell, D. T. Reforms as experiments. American Psychologist, 1969, 24, 409-429.

Campbell, D. T. Methods for the experimenting society. Paper presented at the Meetings of the American Psychological Association, Washington 1971.

Cicirelli, V. G., et al. The impact of Head Start: An evaluation of the effects of Head Start on Children's cognitive and affective development. Vol. 1. Report to the U.S. Office of Economic Opportunity. Columbus, OH: Westinghouse Learning Corporation and Ohio University, 1969.

Cook, T. D., & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. In M. D.

Dunnette (Ed.), Handbook of Industrial and Organizational Psychology. Chicago, IL: Rand McNally, 1976.

Franklin, S., & Thrasher, J. An introduction to program evaluation. New York: Wiley, 1976.

Gergen, K. On taking dialectics seriously. Personality and Social Psychology Bulletin, 1977, 3, 714-718.

Guttentag, M., & Streuning, E. L. Handbook of Evaluation Research. Beverly Hills, CA: Sage Publications, 1975.

Kelling, G. L., Pate, T., Dieckman, D., & Brown, C. E. The Kansas City Preventive Patrol Experiment. Washington, D.C.: Police Foundation, 1974.

Kershaw, D. N. Issues in income maintenance experimentation. In P. H. Rossi and W. Williams (Eds.), Evaluating social programs: Theory, practice and politics. New York: Seminar Press, 1972.

Kruglanski, A. W. The human subject in the psychology experiment: Fact and artifact. In L. Berkowitz (Ed.), Advances in experimental social psychology, Vol. 8. New York: Academic Press, 1975.

Lewin, K. Resolving social conflicts. New York: Harper, 1948.

McCord, J. A thirty-year follow-up of treatment effects. American Psychologist, 1978, 33, 284-289.

McDill, E. L., & McDill, M. S., & Sprehe, J. T. Evaluation in practice: Compensatory education. In P. H. Rossi and W. Williams (Eds.), Evaluating social programs: Theory, practice, and politics. New York: Seminar Press, 1972.

McGuire, W. J. The yin and yang of progress in social psychology: Seven Koan. Journal of Personality and Social Psychology, 1973, 26, 446-456.

Mitroff, I., & Bonoma, T. Psychological assumptions, experimentation, and real world problems: A Critique and an alternate approach to evaluation. Evaluation Quarterly, 1978, 2, 235-260.

Newhouse, J. P. A design for a health insurance experiment. Inquiry, 1974, 11, 5-27.

Perloff, R., Perloff, E., & Sussna, E. Program evaluation. In M. R. Rosenzweig and L. W. Porter (Eds.), Annual Review of Psychology, Vol. 27. Palo Alto, CA: Annual Reviews, 1976.

Rosenthal, R. & Rosnow, R. L. The volunteer subject. In R. Rosenthal and R. L. Rosnow (Eds.), Artifact in behavioral research. New York: Academic Press, 1969.

Rossi, P. & Lyall, C. Reforming public welfare: A critique of the negative income tax experiment. New York: Russell Sage Foundation, 1976.

Saxe, L., & Fine, M. The many uses of control groups in evaluations. In L. Datta and R. Perloff (Eds.), Improving evaluations. Beverly Hills, CA: Sage, in press.

Schlenker, B. R. Social psychology and science. Journal of Personality and Social Psychology, 1974, 29, 1-15.

Suchman, E. Evaluative research. New York: Russell Sage Foundation, 1967.

Weiss, C. H. Evaluation research: Methods of assessing program effectiveness. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Wertheimer, et al. Psychology and the future. American Psychologist, 1978, 33, 631-647.

Williams, W., & Evans, J. W. The politics of evaluation: The case of Head Start. The Annals of the American Academy of Political and

Social Science, 1969, 385, 118-132.

Wortman, P. M. Evaluation research: A psychological perspective. American Psychologist, 1975, 30, 562-575.